Optimal Sequencing of Presidential Primaries*

Deniz Selman†

Boğaziçi University

10 May 2010

Abstract

The American presidential nomination process consists of a series of elections (primaries) in which states vote at different times. This paper focuses on the problem faced by a political party that wants to choose the *optimal temporal structure* for its primaries. I consider an environment in which a sequential election may generate *voter herding*, and address both when and how the party can benefit from social learning to maximize the probability of selecting the highest quality candidate. By choosing whether to have a sequential election and—if so—the actual sequence in which states vote, the party can control whether momentum effects will be present and guarantee that any voter herding will be ex ante beneficial to the goal of selecting the highest quality candidate. When candidates are expected to have equal loyal support, simultaneous voting outperforms all sequential elections. When one candidate has more loyal support, a sequential election can outperform simultaneous voting; under this condition, voter herding compensates for the loyal voter imbalance. This result is a novel example of a beneficial information cascade, in contrast with the socially inefficient cascades in the standard herding literature. In a sequential election, it is optimal for the states voting early in the primary season to be those which: (i) are smaller, (ii) have fewer loyal voters, (iii) have more informed voters, and (iv) display more voter diversity.

JEL classification: D71, D72, D82, D83.

*This paper was a chapter of my dissertation at the University of Pennsylvania. I am indebted to Antonio Merlo for his guidance and support. I would also like to thank Andrew Postlewaite, Alvaro Sandroni, Philipp Kircher, Mattias Polborn, Eddie Dekel, Steven Callander, Brian Knight, Rebecca Morton, Matias Iaryczower, Aureo De Paula, Navin Kartik, Alessandra Casella, Larbi Alaoui, Irina Telyukova, Shalini Roy, Eleanor Harvill, Nirav Mehta, Antonio Penta, Ben Lester and seminar participants at the University of Pennsylvania for helpful comments. All errors are my own.

†Boğaziçi University, Department of Economics, Natuk Birkan, Bebek, İstanbul, Turkey, TR-34342. email: deniz.selman@boun.edu.tr.

"Leaders in the Republican and Democratic parties are struggling to contain a national calendar revolt...that threatens to shift the 2008 presidential nominating contest into the closing months of 2007. And by all indications, they are losing." - The Washington Post*, 5 May 2007*

1 Introduction

The American presidential nomination process consists of a series of elections (primaries) held across the many states and territories over several months. The sequential nature of primaries has produced presidential nominees that most experts feel would not have won their party's nomination under simultaneous voting. Jimmy Carter's nomination by the Democratic Party in 1976 is a famous example.[1] In an empirical study, Knight & Schiff (2007) estimate that voters in early primary states have up to 20 times more influence in the selection of nominees than later voters.[2]

While the significance of primaries' sequencing is well known and often debated, the existing primary calendar has developed more through tradition, political maneuvering and sheer accident than by central party planning. Still, there is recent evidence of both major political parties attempting to control aspects of their primaries' temporal structure. When Florida and Michigan decided to move their respective primaries to January in the 2008 election, both parties threatened to take away some or all of those states' delegates to prevent the move. In the same election, the Democratic Party allowed both Nevada and South Carolina to hold early primaries in order to increase the influence of minority voters.[3] Despite these and other minor interventions, the following question remains unanswered: *If a party could make all decisions regarding primary dates centrally, what sequence of primaries should it use to select its nominee?*

This paper focuses on the problem faced by a political party that wants to choose the *optimal temporal structure* for its primaries. To identify the best system for the party, I develop a model of a presidential primary with two candidates from the same party and incomplete information. Voters are assumed to either be loyal voters of one of the candidates (meaning they vote for that candidate unconditionally) or uncommitted voters, who have common preferences so that they all wish to vote for the *highest quality candidate*. Nature determines which of the candidates is of highest quality, and all uncommitted voters obtain private signals regarding candidate quality. Results from state primaries are observed by all voters as they occur. Uncommitted voters are Bayesian: they vote for the candidate who is best with the highest posterior probability at the time they must vote.[4] The party's nominee is determined by majority rule over all states. Given this framework, the political party chooses the temporal structure of the election to maximize the probability of electing the highest quality candidate. In doing so, the party faces two obstacles: first, it is uncertain which candidate is best; second, it does not know the exact number of loyal voters and uncommitted voters in the party's electorate.

The driving force behind the model is the response of uncommitted voters to different systems of primaries. While loyal voter behavior is fixed, uncommitted voters' behavior depends on the temporal structure of the

[1] Bartels (1988) dedicates an entire chapter, "The Case of Jimmy Carter", to describing how Carter gained crucial support and momentum as a direct result of surprising victories in early primaries.

[2] Using voters' reactions to daily opinion polls in several states in the 2004 Democratic primaries, Knight & Schiff (2007) find that early primary results influence later voters' opinions. In particular, late voters adjust their opinions relative to their expectations before the early primaries. Such a reaction by late voters corresponds to theoretical results in both Ali & Kartik (2007) and this paper.

[3] The Rules Committee of the Democratic National Committee specifically picked Nevada to respond to complaints that western states and minority voters (14% of Nevada's voters are Hispanic) were underrepresented in early primaries, making it the first ever state whose primary date was directly determined by a national party.

[4] Ali & Kartik (2007) refer to this voting behavior as *posterior-based voting.*

primaries. In a simultaneous election, uncommitted voters have only their own private information to use when voting, so all dispersed information aggregates during the election. In contrast, a sequential election may generate *voter herding* amongst uncommitted voters. As in the literature on observational social learning (e.g. Banerjee 1992, Bikhchandani et al. 1992), voters in later periods may choose to ignore their private signals in favor of following the lead of earlier voters. This herding results in a loss of information while emphasizing the importance of early states' results.

I address when and how the party can *benefit* from voter herding–in spite of the loss of valuable information–to maximize the probability of selecting the highest quality candidate. By choosing whether to have a sequential election and—if so—the actual sequence in which states vote, the party can control whether momentum effects will be present and guarantee that any herding will be ex ante beneficial to the goal of selecting the highest quality candidate.

My findings are as follows: first, I show that a simultaneous election outperforms all sequential mechanisms when both candidates are expected to have equal loyal support and voters across all states are ex ante homogeneous. In this case, the political party does best by allowing all private information from uncommitted voters to aggregate without momentum effects.

In contrast, if one candidate has more expected loyal support than the other, a sequential election outperforms simultaneous voting if there is an adequate combination of loyal voter imbalance and low quality of private information. Moreover, the party necessarily benefits from voter herding with even a very small loyal voter imbalance when the population is sufficiently large. This result contrasts that of Banerjee (1992), who finds that herding "may be (and for a large enough population, *will* be) *inefficient* in the ex ante welfare sense."[5] In this paper, voter herding in a sequential election may be optimal when there is a loyal voter imbalance, and *will* be *optimal* for a large enough population.

The main reason for the contrast with the herding literature is that voter herding in my model serves to compensate for an imbalance in loyal support for the two candidates. By choosing a sequential election, the party allows for social learning among uncommitted voters to take place across states. The herding in later states can generate enough votes for the high quality candidate to overcome the imbalance of loyal voters. Under the same conditions, simultaneous voting may not induce a strong enough result to overcome this imbalance.

As in the simple herding literature, social learning entails the risk that voters may herd to the wrong candidate. The party takes on this risk in a sequential election; the quality of information from early states alone (compared to all states) provides a less accurate estimate of which candidate is best. If candidates have equal expected loyal support, there is no initial imbalance which needs to be compensated for, and so the party is best off not taking the risk of negative herding. Choosing the less risky primary system—a simultaneous election—is therefore optimal in this case; all available information is thus utilized to obtain the most precise estimate of which candidate is best.

I then analyze the optimal sequence of states given that a sequential election is preferred by the party. States differ in fraction of informed voters, fraction of loyal voters, and the degree to which voters within the state differ (voter diversity). It is by selectively increasing the effective importance of results from early states that the party can best benefit from herd behavior. I show that, holding each of the other characteristics constant, it is optimal for the states voting early in the primary season to be those which: (i) are smaller, (ii) have fewer loyal voters, (iii) have more informed voters, and (iv) display more voter diversity.

The paper proceeds as follows: Section 2 provides a short history of primaries. Section 3 reviews the related literature. The model is described in Section 4, and optimal voting behavior is characterized in Section 5.

[5] Welch (1992) and Bikhchandani et al (1992) arrive at similar results in variations of the simple herding model.

Section 6 contains the main results regarding when a sequential election outperforms simultaneous voting. The optimal voting sequence in a sequential election is analyzed in Section 7. Section 8 provides a discussion regarding the temporal structure of primaries in recent elections, including some empirical observations of patterns in calendar changes in the last three election cycles. Section 9 concludes. Proofs can be found in the Appendix.

2 A brief history of primaries

Neither the U.S. Constitution nor federal law constrains the method by which political parties choose their presidential nominees.[6] Candidates were chosen by caucuses in Congress until the 1830s, when a system of national party conventions developed to nominate candidates. These conventions were attended by delegates chosen by state and local level party insiders. Candidates were chosen in the proverbial "smoke-filled back rooms" by professional party politicians bargaining for delegates.

In 1901, Florida became the first state to allow parties to choose their delegates with a primary election. Four years later, a Wisconsin statute was the first to *require* that parties do so. Oregon went one step further in 1910, making it mandatory for elected delegates to vote for the actual candidate for whom primary voters had expressed preference. Between 1920 and 1968, however, an average of only 15 states held primaries, and only 40% of all delegates were chosen or bound by primary voters. While the nominating system in those years was in principle a mix of inside strategies and popularity among voters, the system still relied very heavily on inside politics rather than primaries.

The events surrounding the 1968 Democratic primary ultimately brought the end of the mixed system. The anti-war candidate Eugene McCarthy won several early primaries while latecomer Robert Kennedy was successful in later ones. After Kennedy was assassinated, however, party insiders succeeded in nominating Hubert Humphrey, who had not competed in any primaries at all. While Humphrey went on to lose the general election, McCarthy supporters felt their candidate had been wronged, blaming irregularities and secrecy in the system. Riots outside the convention in Chicago reflected the feelings of such voters, and led to the McGovern-Fraser Reforms, which explicitly rejected the notion that party insiders would run the nomination process. The Republicans quickly followed suit, and winning primaries has been the key to winning both parties' nominations since 1972.

The current temporal structure of primaries has risen mainly through coincidence and state-level political maneuvering. In 1916, New Hampshire chose its primary date to be on its town meeting day, scheduled for March to avoid the ensuing "mud season" during which roads were closed and farmers could not travel into town. It is for this reason that New Hampshire has its "first-in-the-nation" presidential primary status. The other famously early state, Iowa, had its caucus moved up in 1972 by an operative for George McGovern as part of the McGovern-Fraser Reforms.

Other states have tried to increase their influence by holding earlier primaries. The current trend of "front-loading", where multiple states move their primaries earlier in the year, began in 1988 when a group of southern states moved their dates to increase regional influence. New Hampshire and Iowa later responded with laws requiring themselves to go first. A political struggle over the primary calendar has followed, both among states and between states and the national parties. As a result of this political maneuvering, there has been significant movement in states' primary dates in recent elections. Please see Section 8 for a more detailed discussion regarding recent changes in the primary calendar and how they relate to the theoretical

[6]See Part 1 of DiClerico & Davis (2000) and Chapter 1 of Bartels (1988) for more detailed summaries of the history of the nomination process.

results in this paper.

3 Related literature

The literature on observational social learning–which I will also refer to as the pure herding literature–originated with Banerjee (1992), Welch (1992) and Bikhchandani et al. (1992), the last of which includes presidential primaries as an example to which their model applies.[7] In these papers, agents with private information and common values sequentially make a decision over a set of alternatives of uncertain value. All decisions are made public so that agents may choose to ignore their private signals and vote based on the observed actions of agents who decide before them.

There are several papers which analyze the temporal structure of elections. Dekel & Piccione (2000) show that the temporal structure does not matter since voting equilibria of sequential elections are the same as those of simultaneous elections. Morton & Williams (1999, 2001) examine herding with multiple candidates and voters with varying risk aversion. Witt (1997) and Fey (2000) consider common-value environments in which herding may occur in a sequential election. More recently, Callander (2007) considers a sequential election in which voters get utility from both conforming and voting for the likely best candidate. Herding results as the desire to conform dominates when the number of voters increases. Iaryczower (2008) considers strategic voting in sequential committees.

A common thread among these papers is that they are positive models of sequential elections with strategic voters. In this paper, I focus on the *normative* question of which system of primaries is optimal from the party's point of view. In doing so, I restrict attention to sincere voting to simplify the analysis. Even when strategic considerations are present in this environment, sincere voting–for the candidate with the highest posterior probability of being best–can be supported as an equilibrium. Ali & Kartik (2007) refer to this type of voting as *posterior-based.* They consider an environment similar to this paper, in which voters are either partisans or neutral. Neutral voters in early periods are aware that neutral voters in later periods may use their vote as information, and vote strategically to maximize the probability the best candidate is elected. Posterior-based voting is shown to be an equilibrium of the voting game.

To the best of my knowledge, Klumpp & Polborn (2006) provide the only mainly normative analysis of the temporal structure of presidential primaries. In a key contrast to this paper, they focus exclusively on candidates' reactions to primary results by endogenizing campaigning behavior. They show that the winning candidate from the first period has more incentive to campaign harder in later periods than the losing candidate, leading to a momentum effect. State primary results are random variables which depend on campaign expenditures in that state only. "Voters" are unaffected by all other factors, including other states' primary results. Klumpp & Polborn (2006) find that a sequential system results in lower levels of advertising expenditures than does a simultaneous election, and is also more likely to select the more effective campaigner.

4 The model

There is a primary election with two candidates ($c = A, B$) from the same political party. Nature determines which candidate $c^{HQ} \in \{A, B\}$ is of highest quality, placing probability 1/2 on each. The election is contested in a total of $S \in \mathbb{N}$ states. Each state $s \in \mathcal{S} = \{1, \ldots, S\}$ holds an election which determines $d_s \in \mathbb{N}$ delegates, assumed to be proportional to the state's population. Let v_s^c be the number of votes cast for candidate c

[7]See Bikhchandani et al (1992) pg. 1010.

in state s. Delegates are assigned by proportional representation.[8] The number of delegates d_s^A received by candidate A is given by

$$d_s^A = \left(\frac{v_s^A}{v_s^A + v_s^B} \right) \cdot d_s.$$

The candidate receiving the majority of the $\sum_{s \in \mathcal{S}} d_s$ delegates wins the party's nomination.

Voter preferences are as follows. First, voting is costless for all voters, so all voters vote in the primary election. Each state contains three types of voters: (i) *loyal voters* of candidate A, (ii) *loyal voters* of candidate B, and (iii) *uncommitted voters.* Loyal voters are characterized by a rigid preference for one candidate. A loyal voter of candidate c gets utility one from voting for c and zero otherwise. Uncommitted voters get utility one from voting for c^{HQ} and zero otherwise. Because the true identity of c^{HQ} is never observed, the relevant object to analyze for uncommitted voters is their *expected* utility at the time they must vote.

4.1 Information structure

To define an uncommitted voter's expected utility, I will first introduce the information structure. An uncommitted voter i in state s receives a private noisy signal $\theta_{i,s} \in \{A, B\}$ regarding which candidate is highest quality, where

$$\Pr\left\{\theta_{i,s} = c^{HQ}\right\} = q_s > 1/2.$$

Depending on the system of primaries ultimately chosen by the political party, elections in different states may be held on different dates. Let $\tau(s) \in \{1, \ldots, T\}$ denote the date on which state s votes. When an election is held in state s, voters in all future states observe the number of votes cast for each candidate in that election, $\left(v_s^A, v_s^B\right)$. All voters voting on date t observe the voting history

$$\mathcal{H}_t = \left\{ \left\{ v_{s'}^A, v_{s'}^B \right\}_{s'} : \tau\left(s'\right) < t \right\}.$$

Uncommitted voters are Bayesian: the expected utility of an uncommitted voter voting for candidate c on date t is the posterior probability (as of date t) that $c = c^{HQ}$. This probability is conditional on private signal $\theta_{i,s}$, private signal quality q_s and observed voting history $\mathcal{H}_{\tau(s)}$. Formally, this expected utility can be written as

$$U_{i,s}(c) = \Pr\left\{c = c^{HQ} \mid \theta_{i,s}, q_s, \mathcal{H}_{\tau(s)}\right\}.$$

The uncommitted voter chooses $c \in \{A, B\}$ to maximize $U_{i,s}(c)$.

4.2 Composition of electorate

The composition of the electorate is determined stochastically. Voters observe only the distribution from which this composition is drawn. This distribution is defined as follows. Each voter in state s receives a preference shock so that she is a loyal voter of candidate A with probability p_s^A, a loyal voter of candidate B with probability p_s^B, and an uncommitted voter with probability $1 - p_s^A - p_s^B$.

The stochastic nature of the electorate is intended to capture idiosyncrasies and various unpredictable aspects of how voters will react to candidates in a primary election. However, it would be unreasonable to assume that preferences and information are determined independently on an individual level. A more reasonable assumption is that there are large blocs of voters–which I will call *voter groups*–whose voting

[8] An extension of the paper which compares proportional representation to a winner-take-all delegate system is in progress. It is useful to note, however, that the assumption of proportional representation does not affect voter herding in this model. Because voters are sincere and have access to complete election results, they would react the same way to a given tally of votes regardless of how that tally maps to a delegate count.

behavior is determined by similar factors. Voters within one voter group are interpreted as those who have similar political views, consume similar media (newspapers, television channels, internet sites, etc.) and have similar reactions to those media, candidate advertising and other sources of political information.

Voter groups are modeled as follows. Each state s has n_s equally sized voter groups. All voters within a voter group receive identical preference and information shocks. Across voter groups within a state, both preference shocks and private signals are independently and identically distributed. Technically, this assumption is equivalent to assuming a smaller number of voters in each state than there actually are. As a result, the initial uncertainty regarding which is the highest quality candidate is preserved after observing early primary results. If preference and information shocks were i.i.d. across individual voters, this uncertainty would be eliminated by the Law of large Numbers.

The notion of voter groups achieves the purpose of capturing both idiosyncrasies and aggregate shocks within a state population. In doing so, it also allows for an explicit representation of *voter diversity* within each state. While voter groups are assumed to be the same size within a state, they may be different sizes across states. Because d_s is proportional to the population of state s, the number of voter groups per delegate, $\lambda_s \equiv n_s/d_s$, is used to denote the voter diversity in state s. Voter diversity λ_s captures how informative the assignment of each delegate in state s is to future voters. The more independent draws n_s are used to determine the composition of a state's electorate, the more informative results from that state will be.

4.3 Political party's problem

The political party is faced with choosing an *optimal temporal structure* for its primaries. Because this structure cannot be specific to candidates in a particular election cycle, I assume that the party does not know the true identity of c^{HQ}. Instead, it has the same (uniform) prior for c^{HQ} as uncommitted voters. Additionally, in the absence of candidate identities, the party also cannot observe the exact number of loyal voters and uncommitted voters in the electorate. The party observes only the distribution, described above, from which this composition is drawn.

I can now formalize the political party's problem. An **ordered partition** O of the set of states $\mathcal{S}$ is given by

$$O = \mathcal{S}_1, \ldots, \mathcal{S}_T,$$

where $\mathcal{S}_1, \ldots \mathcal{S}_T$ (with $1 \leq T \leq S$) are non-empty subsets of $\mathcal{S}$ such that

$$\bigcup_{t=1}^{T} \mathcal{S}_t = \mathcal{S}.$$

Let $\mathcal{O}(\mathcal{S})$ be the set of ordered partitions of $\mathcal{S}$. Then, each ordered partition $O \in \mathcal{O}(\mathcal{S})$ represents one of the possible temporal structures for primaries which the party can choose, where

$$\mathcal{S}_t = \{s \in \mathcal{S} : \tau(s) = t\}$$

denotes the set of states whose elections are held on date t. Let $W(O)$ denote the realized winner of an election using structure O. The party's problem is given by

$$\max_{O \in \mathcal{O}(\mathcal{S})} \Pr\left\{W(O) = c^{HQ} \mid \{q_s, p_s^A, p_s^B\}_{s \in \mathcal{S}}\right\}.$$

5 Voting behavior

Loyal voter voting behavior is fixed. All loyal voters always vote for their own candidate regardless of the temporal structure of primaries or when in the sequence they vote. The rest of this section analyzes the voting behavior of uncommitted voters.

Uncommitted voters in the first period have only private information on which to base their vote, since $\mathcal{H}_1 = \varnothing$. They hence all vote in accordance with private signal $\theta_{i,s}$. To characterize optimal voting behavior for uncommitted voters in periods $t \geq 2$, I introduce the following notation. Let γ_s^c represent the probability that a randomly selected voter in state s, with $\tau(s) = 1$, votes for c given that $c = c^{HQ}$. This probability is simply the sum of the probability that the voter is a loyal voter of candidate c plus the probability that the voter is an uncommitted voter who received the correct private signal:

$$\gamma_s^c = p_s^c + \left(1 - p_s^A - p_s^B\right) q_s.$$

Because the election in state s is simply the result of n_s independent draws of such a voter, the number of votes for candidate c^{HQ} in state s in the first period is a binomial random variable with n_s trials and probability parameter $\gamma_s^{c^{HQ}}$:

$$v_s^{c^{HQ}} \sim \text{Binomial}\left(n_s, \gamma_s^{c^{HQ}}\right).$$

If state $s = 1$ is the only one to vote in the first period, an uncommitted voter in state $s = 2$ voting in the second period with signal $\theta_{i,2}$ and signal quality q_2 will vote for candidate A iff

$$\begin{aligned} & U_{i,2}(A) > U_{i,2}(B) \\ \Leftrightarrow\ & \Pr\left\{A = c^{HQ} \mid \theta_{i,2}, q_2, \mathcal{H}_2\right\} > \Pr\left\{B = c^{HQ} \mid \theta_{i,2}, q_2, \mathcal{H}_2\right\} \\ \Leftrightarrow\ & f\left(v_1^A \mid c^{HQ} = A\right) \Pr\left\{\theta_{i,2} \mid c^{HQ} = A\right\} > f\left(v_1^A \mid c^{HQ} = B\right) \Pr\left\{\theta_{i,2} \mid c^{HQ} = B\right\}, \end{aligned} \tag{1}$$

where

$$f\left(v_1^A \mid c^{HQ} = A\right) = \binom{n_1}{v_1^A} \left(\gamma_1^A\right)^{v_1^A} \left(1 - \gamma_1^A\right)^{n_1 - v_1^A}$$

and

$$f\left(v_1^A \mid c^{HQ} = B\right) = \binom{n_1}{v_1^A} \left(1 - \gamma_1^B\right)^{v_1^A} \left(\gamma_1^B\right)^{n_1 - v_1^A}$$

are the appropriate binomial density functions. Solving (1) yields the following result:

Lemma 1 *If states $s \in \mathcal{S}_1 = \{1, \ldots, S_1\}$ voting in the first period are characterized by $\left\{p_s^A, p_s^B, q_s, n_s\right\}_{s \in \mathcal{S}_1}$, then the optimal voting decision for a second period uncommitted voter in state s' with private signal $\theta_{i,s'}$ and signal quality $q_{s'}$ is given by*

$$\mathfrak{v}_{i,s'}^* = \begin{cases} A & \textit{for } \prod_{s \in \mathcal{S}_1} \left(\frac{f\left(v_s^A \mid c^{HQ} = A\right)}{f\left(v_s^A \mid c^{HQ} = B\right)}\right) > \frac{q_{s'}}{1 - q_{s'}} \\ \theta_{i,s'} & \textit{for } \frac{1 - q_{s'}}{q_{s'}} < \prod_{s \in \mathcal{S}_1} \left(\frac{f\left(v_s^A \mid c^{HQ} = A\right)}{f\left(v_s^A \mid c^{HQ} = B\right)}\right) < \frac{q_{s'}}{1 - q_{s'}} \\ B & \textit{for } \prod_{s \in \mathcal{S}_1} \left(\frac{f\left(v_s^A \mid c^{HQ} = A\right)}{f\left(v_s^A \mid c^{HQ} = B\right)}\right) < \frac{1 - q_{s'}}{q_{s'}}, \end{cases} \tag{2}$$

where

$$f\left(v_s^A \mid c^{HQ} = A\right) = \binom{n_s}{v_s^A} \left(\gamma_s^A\right)^{v_s^A} \left(1 - \gamma_s^A\right)^{n_s - v_s^A}$$

and

$$f\left(v_s^A \mid c^{HQ}=B\right)=\binom{n_s}{v_s^A}\left(1-\gamma_s^B\right)^{v_s^A}\left(\gamma_s^B\right)^{n_s-v_s^A}$$

are the appropriate binomial density functions. If n_1 voter groups all characterized by (p_1^A, p_1^B, q_1) vote in the first period, with v_1^A of them voting for candidate A, then (2) can be simplified to

$$\mathfrak{v}_{i,s'}^{*}=\begin{cases} A & \text{for } v_1^A > \Omega^A(n_1) \\ \\ \theta_{i,s'} & \text{for } \Omega^B(n_1) \le v_1^A \le \Omega^A(n_1) \\ \\ B & \text{for } v_1^A < \Omega^B(n_1), \end{cases}$$

where

$$\begin{aligned} \Omega^A(n_1) &\equiv \Omega^A(n_1; q_{s'}) = \frac{n_1 \ln\left(\frac{\gamma_1^B}{1-\gamma_1^A}\right)+\ln\left(\frac{q_{s'}}{1-q_{s'}}\right)}{\ln\left(\frac{\gamma_1^A}{1-\gamma_1^B}\right)+\ln\left(\frac{\gamma_1^B}{1-\gamma_1^A}\right)}, \\ \Omega^B(n_1) &\equiv \Omega^B(n_1; q_{s'}) = \frac{n_1 \ln\left(\frac{\gamma_1^B}{1-\gamma_1^A}\right)-\ln\left(\frac{q_{s'}}{1-q_{s'}}\right)}{\ln\left(\frac{\gamma_1^A}{1-\gamma_1^B}\right)+\ln\left(\frac{\gamma_1^B}{1-\gamma_1^A}\right)} < \Omega^A(n_1). \end{aligned}$$

The intuition behind lemma 1 is straightforward. When candidate A receives more than $\Omega^A(n_1)$ votes in period 1, even uncommitted voters who receive a private signal of $\theta_{i,s'} = B$ in period 2 believe with probability greater than 1/2 that $c^{HQ} = A$. The indifference curve for such voters is $\Omega^A(n_1)$. Symmetrically, $\Omega^B(n_1)$ is the indifference curve for those uncommitted voters who receive a signal of $\theta_{i,s'} = A$. For all results such that candidate A receives fewer than $\Omega^B(n_1)$ votes, all uncommitted voters believe with probability greater than 1/2 that $c^{HQ} = B$.

Figures 1-4 provide graphs of $\Omega^A(n_1)$ and $\Omega^B(n_1)$ under different conditions on the parameters p_1^A, p_1^B and q_2. There are two key features to this sensitivity analysis. First, when $p_1^A > p_1^B$, candidate A must receive more first period votes to induce second periods voters to herd to him than does candidate B. This is because second period voters know that candidate A is expected to have more loyal support and attribute a portion of his vote total to this fact. Second, when second period voters have a higher signal quality $q_{s'}$, they are less likely to herd at all. The decision to herd rather than vote with one's private information can be interpreted as trusting public information (in this case, the results of the first state's primary) more than one's private information. When the quality of second period voters' private information increases, so too does the threshold of how many first period votes a candidate must receive in order to convince uncommitted voters receiving the opposite signal to vote for him.

Lemma 1 characterizes herding in the second period only. Herding behavior in future periods is addressed by the next result.

Let n_t denote the number of votes cast in period t, and let v_t^A be the number of them cast for candidate A.

Lemma 2 *Suppose $q_s \ge q_{s'}$ for all s, s' such that $\tau(s) < \tau(s')$, and let $q_{t^{\max}} \equiv \max\{q_s : \tau(s) = t\}$. Let*

$$\widehat{t} \equiv \min\left\{t : \prod_{s:\tau(s)<t}\left(\frac{f\left(v_s^A \mid c^{HQ}=A\right)}{f\left(v_s^A \mid c^{HQ}=B\right)}\right) \notin \left[\frac{1-q_{t^{\max}}}{q_{t^{\max}}}, \frac{q_{t^{\max}}}{1-q_{t^{\max}}}\right]\right\}$$

denote the first period in which all uncommitted voters herd to one candidate. Then, uncommitted voters in all states s with $\tau(s) > \widehat{t}$ will herd to the same candidate as well.

Figure 1

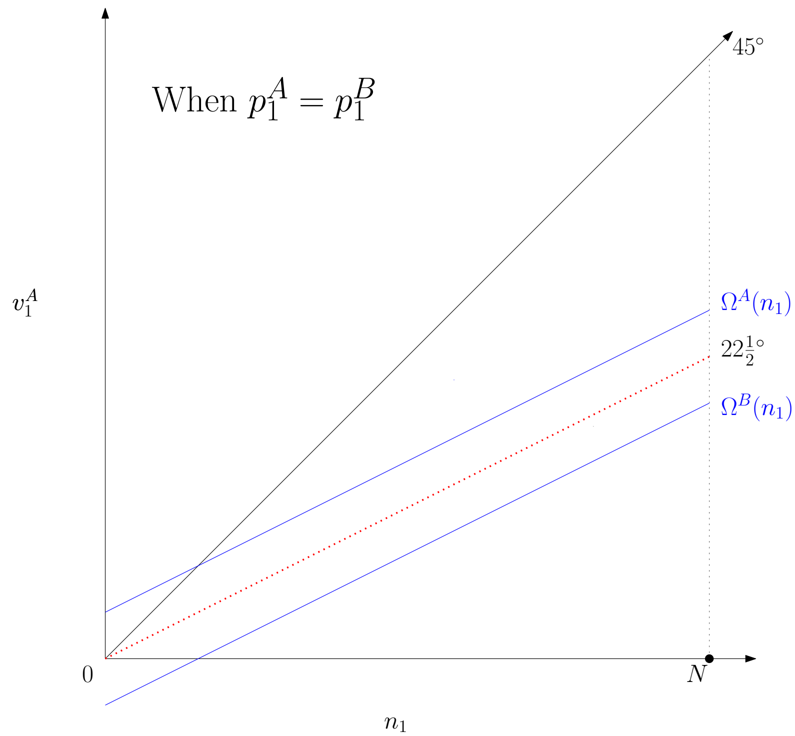


Figure 2

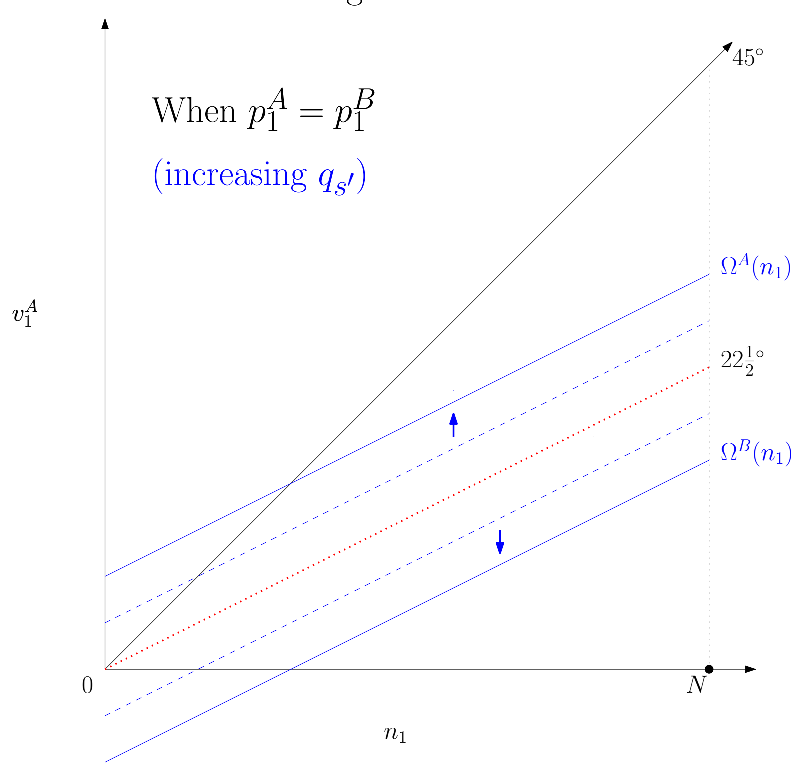


Figure 3

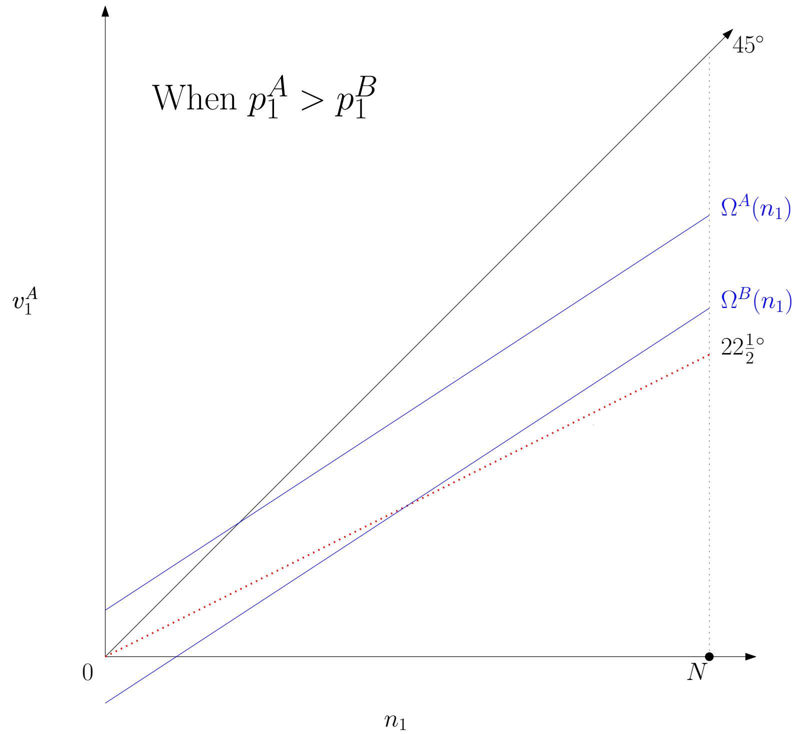
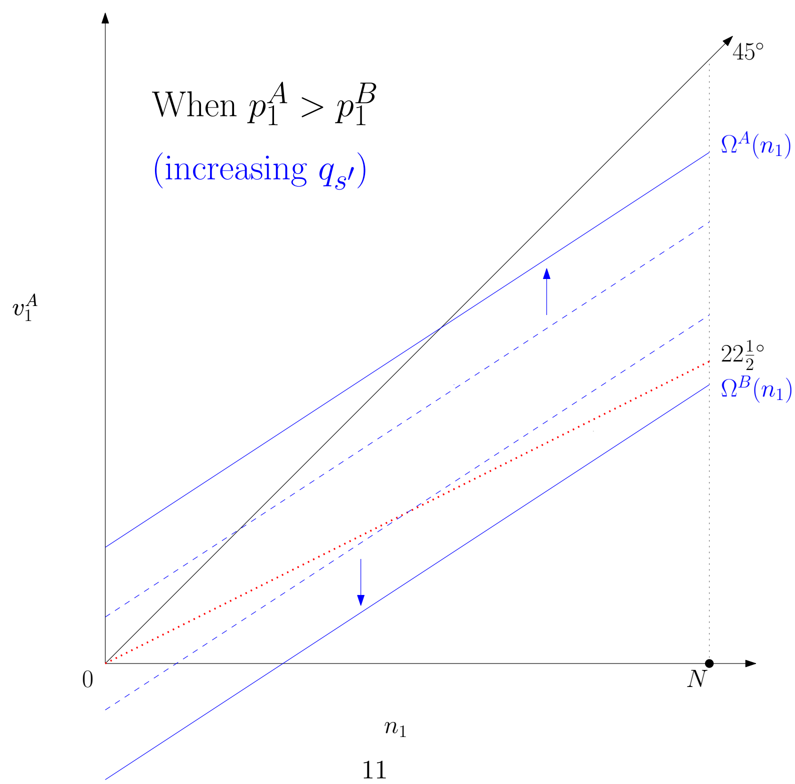


Figure 4



Lemma 2 simply states that once herding occurs, all future uncommitted voters will join the herd if they do not have better private information than those who voted before them. Once herding begins, results from all periods with full herding are ignored by future voters because they know that no private information is being used in those periods. All uncertainty in such periods stems from the number of loyal voters (of the candidate not benefiting from herding), which contains no useful information to uncommitted voters.

6 Sequential vs. simultaneous election

To analyze whether a simultaneous or sequential election is optimal, I will consider the case with homogeneous voters in all states. That is, I assume that $\left(p_s^A, p_s^B q_s, \lambda_s\right) = \left(p^A, p^B, q, \lambda\right)$ for all s, allowing states to still vary in their number of delegates d_s. This simplification is made for tractability of the model and to allow for clearer results. It is important to note, however, that nothing in my preliminary analysis of the heterogeneous model indicated that the basic results do not carry through.

The definition of a *simultaneous* election is staightforward; it is an election in which all states hold their primary on the same date, so that no voter can observe any results from any other states. In this paper, I use the term *sequential* to describe all elections which are not simultaneous. The purpose of this section is to identify when a simultaneous election outperforms all sequential elections. In the seminal herding literature (e.g. Banerjee 1992, Bikhchandani et al. 1992), observational social learning typically results in undesirable information cascades. Agents who could rely on private information instead choose to follow a herd. The resulting equilibrium is ex ante inefficient due to the risk that voters herd to the wrong candidate. In contrast, simultaneous voting does not allow for such an information cascade and therefore does not lead to this kind of inefficient result.

I find that sequential elections can outperform simultaneous voting under a wide range of conditions. This result is in contrast with the prevailing notion that herding is inefficient. The main reason for this contrast is that voter herding in my model can serve to compensate for an imbalance in loyal support for the two candidates. In fact, one necessary condition for a sequential election to be optimal is for candidates to have different numbers of expected loyal voters $\left(p^A \neq p^B\right)$. Throughout this section, let $N \equiv \sum_{s \in \mathcal{S}} n_s$ denote the total number of voter groups. Also, in a slight abuse of notation, let $n_t \equiv \sum_{s:\tau(s)=t} n_s$ denote the number of voter groups voting on date t.

6.1 The case with equal loyal support

When both candidates are expected to have equal loyal support, I find that a simultaneous election outperforms all sequential elections. In fact, I prove an even stronger result. When calculating the probability of electing c^{HQ} under a sequential election, the political party must integrate out all possible results from early periods. Theorem 1 states that even if the party were given the option of first observing the first period voting result and then deciding whether or not to "reveal" the result to future voters, it would not reveal the result regardless of what it was. Note here that not revealing the result to future voters is informationally equivalent to holding a simultaneous election.

Theorem 1 *Suppose $\left(p_s^A, p_s^B q_s, \lambda_s\right) = \left(p^A, p^B, q, \lambda\right)$ for all s, and that $p^A = p^B$. Let $N \equiv \sum_{s \in \mathcal{S}} n_s$. For any N and any $q > 1/2$, and for any possible first period pair $\left(n_1, v_1^A\right)$ of number of votes n_1 of which v_1^A are cast for candidate A, the probability of selecting the highest quality candidate is at least as high when $\left(n_1, v_1^A\right)$ is not revealed to the remaining $N - n_1$ voters as when it is revealed. For parameter values such that $\left(n_1, v_1^A\right)$ does not secure the nomination for either candidate but does induce herding in the second period, revealing* *the tally results in a strictly lower probability of selecting the higher quality candidate than not revealing. For all other parameter values, revealing the tally has no effect.*

Together with lemma 2, theorem 1 results in an immediate corollary.

Corollary 1 *When voters across states are ex ante homogeneous and $p^A = p^B$, a simultaneous election outperforms any sequential election.*

The result that preventing voter herding from occurring is optimal is similar to the findings in the herding literature that herding is inefficient in an ex ante social welfare sense. Indeed, the uncommitted voters in this model behave in much the same way as agents in that literature. When loyal voters are expected to cancel each other out in number, the election is effectively decided by these uncommitted voters. In such a case, the probability of selecting the highest quality candidate coincides with the probability of correctly identifying that candidate through information revelation from uncommitted voters. This probability decreases when private information is lost by way of voter herding. As in the herding literature, utilizing all available private information is hence optimal.

6.2 The case with an imbalance of loyal voters

It turns out, however, that the optimality of simultaneous voting does not hold in many cases when there is a loyal voter imbalance. When one candidate–say candidate A–has more loyal support, then a sequential election is optimal if there is an adequate combination of sufficient loyal voter imbalance and low quality of private information. The intuition for this fact can be summarized as one candidate's quality overcoming another candidate's advantage in loyal support when there is voter herding.

A numerical example may help illustrate the phenomenon. Abstracting from the stochastic composition of the electorate, suppose $p^A = 1/4$ and $p^B = 0$, so uncommitted voters make up 3/4 of the electorate. For candidate B to win the election, he needs to receive at least 2/3 of the uncommitted votes. If $q < 2/3$, however, fewer than 2/3 of the uncommitted voters vote for B in a simultaneous election even if he is the high quality candidate. As a result, candidate A will win the nomination simply because he has more loyal voters. This is where a sequential election can help the party get the best candidate selected. A sequential election induces voter herding which serves to correlate uncommitted voters to all vote for the highest quality candidate in later voting periods. Rather than only a fraction $q < 2/3$ voting for the best candidate, voters herding to the best candidate can ensure the best candidate receives more than the necessary 2/3 of uncommitted votes to win the nomination.

While a sequential election involves the risk that later voters will herd to the wrong candidate, this risk is worth taking for the party in cases such as the above example when a simultaneous election is likely to simply select the candidate with more loyal voters. This result is a novel example of a beneficial information cascade, in contrast with the socially inefficient cascades in the standard herding literature. Theorem 2 helps to formalize the result, stating that when $p^A > p^B$, a sufficiently large number of total voter groups and a sufficiently low quality of private information ensure that a sequential election is optimal.

Theorem 2 *Suppose $\left(p_s^A, p_s^B q_s, \lambda_s\right) = \left(p^A, p^B, q, \lambda\right)$ for all s, and that $p^A > p^B$. Then:*

1. *For any $q > 1/2$, there exists $N^*(q) > 0$ such that a sequential election outperforms simultaneous voting iff $N > N^*(q)$.*
2. *For any $N > \frac{1-2\left(p^A\right)^2+p^A-p^B+2p^Ap^B}{\left(1-p^A-p^B\right)\left(p^A-p^B\right)}$, there exists $q^*(N) > 1/2$ such that a sequential election outperforms simultaneous voting iff $q < q^*(N)$.*

It should be noted that as the loyal voter imbalance $p^A - p^B$ increases, $N^*(q)$ decreases for all q and $q^*(N)$ increases for all N. That is, the greater the loyal voter imbalance, the larger the set of other parameters for which a sequential election is optimal. The reason low levels of q result in a sequential election being optimal is illustrated in the numerical example above. If $q > 2/3$ in the example, then the party would not need to take the risk of voters herding to the wrong candidate in order to get c^{HQ} nominated. The reason a large number of total voter groups leads to an optimal sequential election is as follows. Because a certain number of independent draws is needed to provide a decent estimate of which candidate is c^{HQ}, an election with a small total number N of voter groups would not have enough voters left over to benefit from the early voters and make a difference in the result by herding. At the opposite extreme, as the number of voter groups grows arbitrarily large, the probability of the first period estimate being correct approaches one, while an arbitrarily large number of voters are still left over to herd to c^{HQ} and ensure he gets nominated.

As discussed in Section 1, the result that voter herding can be optimal contrasts the simple herding literature's result that herding is inefficient in an ex ante welfare sense. The main reason is the loyal voter imbalance for which herding acts as compensation. A second reason is that the political party is concerned only with which candidate wins the nomination, and not the margin of victory or defeat. While a sequential election increases the probability of a lopsided victory by the wrong candidate, it can (at the same time) still decrease the probability that the wrong candidate wins at all. The measure of ex ante welfare in Banerjee (1992) is simply the expected fraction of people who make the correct choice, so a lopsided defeat of the high quality candidate would correspond to a lower welfare level than would a close defeat. In this paper, a defeat by any margin of the high quality candidate yields no return for the party. In an ex ante sense, the simple probability of selecting the high quality candidate is the significant measure of success.

To more precisely characterize when a sequential election is best, consider a two-period party problem with homogeneous voters. Abstracting from exact state sizes, let $n_1 \in [1, N]$ be the party's choice variable of how many voter groups will vote in period 1, with $n_2 = N - n_1$ voters in period 2. Let $V(n_1)$ denote the party's objective function: the probability of electing c^{HQ} when n_1 votes are cast in period 1. Lemma 3 states that to see if any sequential election exists that outperforms simultaneous voting, it is enough to check if the sequential election with the highest number of possible voters in the first period (while still allowing the possibility of relevant herding) outperforms simultaneous voting.

Lemma 3 *Suppose $\left(p_s^A, p_s^B q_s, \lambda_s\right) = \left(p^A, p^B, q, \lambda\right)$ for all s, and that $p^A > p^B$. Let $\overline{n}_1 \leq N$ denote that which satisfies*

$$\Omega^B(\overline{n}_1) = \overline{n}_1 - \frac{N}{2}$$

if one exists and $\overline{n}_1 = N - 1$ if one does not. (Then $\overline{n}_1$ is the maximum number of first period voters such that future herding can affect the election result.) There exists a sequential election which outperforms simultaneous voting iff $V(Floor(\overline{n}_1)) > V(N)$.

Figure 5 shows the case with $\overline{n}_1 = N - 1$. Note that $\overline{n}_1 = N - 1$ holds whenever $p^A = p^B$. Figure 6 shows the case in which $\overline{n}_1 < N - 1$. Lemma 3 makes it significantly easier to analytically determine the parameter values for which a sequential election is optimal. If $\overline{n}_1$ voters have voted in period 1 and the result is potential *relevant* herding (meaning that neither candidate has secured the nomination when herding begins), then the trailing candidate (necessarily the one with fewer loyal voters, candidate B) must receive all remaining votes to win the nomination. Conditional on this first period result, if the probability that $c^{HQ} = B$ is not sufficiently high that the party prefers herding to occur, Lemma 3 states that the party cannot want herding to occur after any number of first period voters.

Figure 5

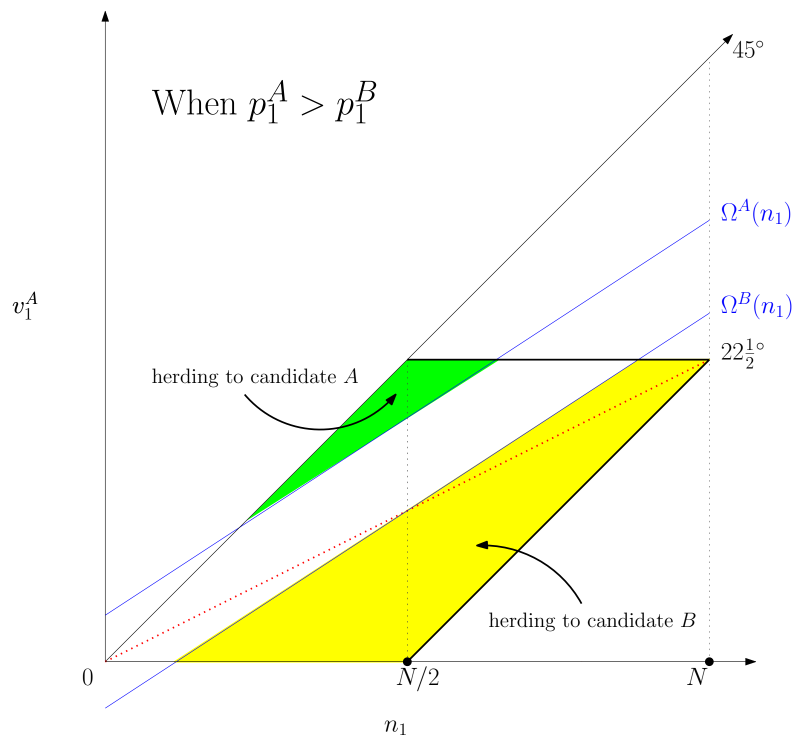
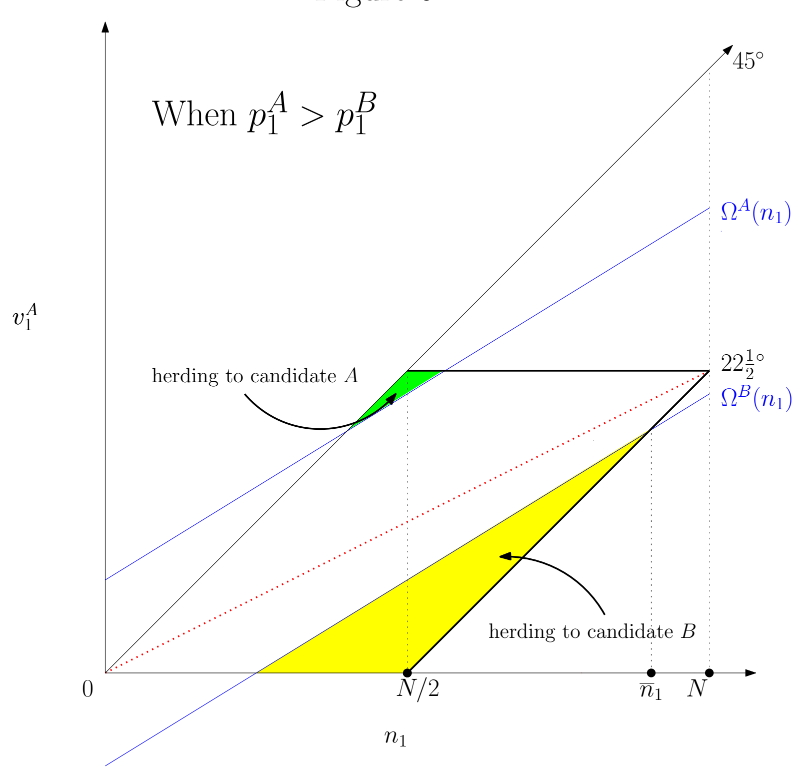


Figure 6



7 Optimal sequencing

Given that the party uses a sequential election, it must also determine what actual sequence of states is best. In this section, I allow states voting in different periods to have different characteristics. Suppose for simplicity that there are two states (s, s') which are to vote in sequential periods. Theorem 3 provides results regarding which state should vote first when the states differ along one dimension at a time.

Theorem 3 *1. Suppose* $\left(p_s^A, p_s^B\right) = \left(p_{s'}^A, p_{s'}^B\right), q_s = q_{s'}$ *and* $d_s > d_{s'}$. *If* $n_s = n_{s'}$, *then it is optimal for state* s' *to vote first. If* $\lambda_s = \lambda_{s'}$, *then the optimal sequence depends on the exact parameters.*

2. Suppose $d_s = d_{s'}, q_s = q_{s'}, \lambda_s = \lambda_{s'}$ *and* $p_s^A + p_s^B > p_{s'}^A + p_{s'}^B$. *If* $\left(p_s^A / p_s^B\right) = \left(p_{s'}^A / p_{s'}^B\right)$, *then it is optimal for state* s' *to vote first. If* $\left(p_s^A / p_s^B\right) \neq \left(p_{s'}^A / p_{s'}^B\right)$, *then the optimal sequence depends on the exact parameters.*[9]

3. If $d_s = d_{s'}, \left(p_s^A, p_s^B\right) = \left(p_{s'}^A, p_{s'}^B\right), \lambda_s = \lambda_{s'}$ *and* $q_s > q_{s'}$, *then it is optimal for state* s *to vote first.*

4. If $d_s = d_{s'}, \left(p_s^A, p_s^B\right) = \left(p_{s'}^A, p_{s'}^B\right), q_s = q_{s'}$ *and* $\lambda_s > \lambda_{s'}$, *then it is optimal for state* s *to vote first.*

The intuition behind these results is as follows. When having *fewer* voters vote in first period, the party is accepting the following trade-off: there is a lower probability that c^{HQ} will be correctly identified by first period voting results, but more second period voters who will herd, and hence a higher probability of *pivotal* herding. As long as states carry the same level of informativeness, it is better if they are smaller so that more voters remain for the second period. Fewer loyal voters and more informed voters both make an early state's results more informative to future states. Hence both are optimal. Higher voter diversity makes a state's results more informative as well, and hence is optimal in early states.

8 Discussion

As mentioned in Section 2, the primary calendar has seen significant changes in recent elections. These changes pose the following questions: *Is there a systematic pattern when comparing early states to late states? What demographic characteristics are correlated with states' primary dates?*

To begin answering these questions, one can choose from a wide range of state level data. In this section I present some data on state-level characteristics that I conjecture are correlated with the theoretical parameters of my model. The goal of this section is *not* to rationalize recent movements in primary dates, nor even to claim any causal connection between the movements and the state characteristics presented. Indeed, there is very little reason to believe that such a causal relationship exists. As mentioned in Sections 1 and 2, political maneuvering and coincidences independent of any notion of optimality have been responsible for most of the changes. However, one can still observe, on a casual level, whether changes in the recent election cycles have been going toward the optimal sequence found in Section 7 or away from such a sequence. Recall that in a sequential election, it is optimal for the states voting early in the primary season to be those which have fewer loyal voters, have more informed voters and display more voter diversity.

Table 1 contains some observed patterns from the last three election cycles. For each election (2000, 2004, 2008), states are divided into two groups based on when the Democratic primary was held.[10] States with primaries on or before "Super Tuesday" are categorized as early, and all other states are categorized

[9] Even when $\left(p_s^A / p_s^B\right) \neq \left(p_{s'}^A / p_{s'}^B\right)$, it is optimal for state s' to vote first for all but a very small subset of the parameter space.

[10] In most states, the Democratic and Republican primaries are held on the same day. For this table, using the date of the Republican primary results in nearly identical figures.

as late.[11] Three state level characteristics are shown. The first, "diversity", comes from the Sullivan index, a measure of population diversity, consisting of variables regarding education, income, occupation, housing ownership, ethnicity and religion. "BA education" is a measure of the fraction of voters in a state with at least a bachelor's degree, while "union" gives union members as a fraction of total workers.

	2000			2004			2008	
	early	late		early	late		early	late
#2008 delegates	1700	2428		2435	1693		2422	1706
diversity	.472	.424	***	.454	.424	*	.444	.436
	(.056)	(.055)		(.059)	(.056)		(.065)	(.052)
BA education	.290	.256	**	.283	.247	**	.269	.266
	(.043)	(.055)		(.055)	(.046)		(.046)	(.063)
union	.145	.113	*	.134	.111		.122	.126
	(.055)	(.055)		(.057)	(.054)		(.059)	(.055)
$p \leq 0.01$:***		$p \leq 0.05$:**			$p \leq 0.10$:*			

Table 1[12]

As can be seen in Table 1, early states displayed significantly more diversity than late states in 2000 according to the Sullivan index. The shifting of the primary calendar reduced this significance in 2004, and eliminated any statistical significance in 2008. Under the conjecture that this notion of diversity correlates with voter diversity (λ_s) in my model, the sequence of states in 2000 matches most closely with the result that more diverse states optimally vote first. Relative to 2000, the 2004 sequence moved away from optimality, while the 2008 calendar moved still further.

A similar decrease in significance from 2000 to 2008 can be seen with the fraction of voters with a bachelor's degree. There are many studies that positively link political knowledge with education level.[13] Assuming such a relationship, the fraction of voters with a bachelor's degree would likely correlate with the quality of private information (q_s) in my model. Recall that states with better private information quality optimally vote first. As was the case with diversity, the 2000 and 2004 primary calendars come closer to matching such an optimal sequencing than the 2008 version, in which early and late states had virtually the same average fraction of voters with a bachelor's degree.

Finally, the fraction of union members is observed to be higher in early states than late states in 2000, with the difference becoming insignificant in 2004 and especially 2008. Union membership is one of many variables one can conjecture correlates with the probability of being a loyal voter as defined by my model. It is undoubtedly more difficult to identify potential loyal voters in a primary election than in a general election. In trying to do so, I attempt to choose voters who are more likely than others to choose a preferred candidate based on a few particular issues for which it is easy to get information. Such voters are less likely to be affected by other states' primary results since these results presumably contain little information regarding the issues important to them. If one assumes union members fit into this class of voters who form a rigid preference for a particular candidate, the 2000 primary calendar display a suboptimal pattern of more loyal voters in early states. The trend along this dimension, unlike the others, is toward a more optimal system since the significance of union membership disappears in 2004 and 2008.

[11] The dates of the last three "Super Tuesdays" were 7 March 2000, 2 March 2004 and 5 February 2008.

[12] Sources: Morgan & Wilson (1990) (diversity), 2006 American Community Survey (BA education), 2006 Current Population Survey, Bureau of Labor Statistics (union).

[13] Campbell et al. (1960) pg. 476 is one example.

9 Conclusion

The American presidential nomination process consists of a series of elections (primaries) in which states vote at different times. This paper has focused on the problem faced by a political party that wants to choose the *optimal temporal structure* for its primaries. Considering an environment in which a sequential election may generate *voter herding*, I have addressed both when and how the party can benefit from social learning to maximize the probability of selecting the highest quality candidate. By choosing whether to have a sequential election and—if so—the actual sequence in which states vote, the party can control whether momentum effects will be present and guarantee that any voter herding will be ex ante beneficial to the goal of selecting the highest quality candidate.

When candidates are expected to have equal loyal support, simultaneous voting outperforms all sequential elections. In a simultaneous election, uncommitted voters cannot rely on any information save their own private signals. Therefore, they vote based on these private signals and the maximum possible amount of information is aggregated in determining the party's nominee. This result is similar to the findings in the standard herding literature (e.g. Banerjee 1992, Bikhchandani et al. 1992) that herding is inefficient in an ex ante social welfare sense.

When one candidate has more loyal support, however, a sequential election can outperform simultaneous voting. Under this condition, voter herding acts as compensation for the loyal voter imbalance. While a simultaneous election may not result in enough votes for the highest quality candidate to overcome an initial loyal voter deficit, a sequential election induces voter herding which serves to correlate uncommitted voters to all vote for the highest quality candidate in later voting periods. While such herding involves the risk that voters will herd to the wrong candidate, it can still increase the probability that the high quality candidate is selected relative to when voting is simultaneous. This result is a novel example of a beneficial information cascade, in contrast with the socially inefficient cascades in the standard herding literature.

The results in Sections 6.1 and 6.2 indicating whether a sequential or simultaneous election is optimal have implications for the welfare effects of political polls as well. Until now, most of the normative literature on political polls has focused on how polls affect turnout. Using the costly voting framework found in Palfrey and Rosenthal (1983), Klor and Winter (2007) show that polls have a positive effect on total welfare when one candidate is expected to have more initial total support. Goeree and Großer (2007) show that polls can reduce welfare by stimulating the minority group to turn out in high enough numbers relative to the majority when an election is close. These papers and others focus on the turnout decision made by voters who already know which candidate they prefer.

This paper provides a groundwork for a normative analysis of political polls which focuses not on turnout but rather the actual voting decision made by voters. The release of poll data can be assumed to affect uncommitted voters in a fashion similar to early primary results. Because uncommitted voters who participate in pre-election polls can choose to use the poll results to change their vote, the trade-off concerning the number of "remaining" voters after the first period is no longer relevant. This fact should only simplify the analysis. Moreover, it is safe to conjecture that the analysis regarding information aggregation will be quite similar to that of this paper.

10 Appendix

Proof of Lemma 1. An uncommitted voter in the second period receiving a private signal of $\theta_{i,2} = A$ with quality q_2 and observing voting history $\mathcal{H}_2 = \left\{v_1^A\right\}$ will vote for candidate A if and only if

$$\begin{aligned}
&U_{i,2}(A) > U_{i,2}(B) \\
&\Leftrightarrow \Pr\left\{c^{HQ} = A \mid \theta_{i,2}, q_2, \mathcal{H}_2\right\} > \Pr\left\{c^{HQ} = B \mid \theta_{i,2}, q_2, \mathcal{H}_2\right\} \\
&\Leftrightarrow \frac{\Pr\left\{v_1^A, \theta_{i,2} = A \mid c^{HQ} = A\right\}\Pr\left\{c^{HQ} = A\right\}}{\Pr\left\{v_1^A, \theta_{i,2} = A\right\}} > \frac{\Pr\left\{v_1^A, \theta_{i,2} = A \mid c^{HQ} = B\right\}\Pr\left\{c^{HQ} = B\right\}}{\Pr\left\{v_1^A, \theta_{i,2} = A\right\}} \\
&\Leftrightarrow \Pr\left\{v_1^A, \theta_{i,2} = A \mid c^{HQ} = A\right\} > \Pr\left\{v_1^A, \theta_{i,2} = A \mid c^{HQ} = A\right\} \\
&\Leftrightarrow \Pr\left\{v_1^A \mid c^{HQ} = A\right\}\underbrace{\Pr\left\{\theta_{i,2} = A \mid c^{HQ} = A\right\}}_{q_2} > \Pr\left\{v_1^A \mid c^{HQ} = B\right\}\underbrace{\Pr\left\{\theta_{i,2} = A \mid c^{HQ} = B\right\}}_{1-q_2} \\
&\Leftrightarrow \frac{\Pr\left\{v_1^A, \theta_{i,2} = A \mid c^{HQ} = A\right\}}{\Pr\left\{v_1^A, \theta_{i,2} = A \mid c^{HQ} = B\right\}} > \frac{1-q_2}{q_2} \\
&\Leftrightarrow \frac{\binom{n_1}{v_1^A}\left(\gamma_1^A\right)^{v_1^A}\left(1-\gamma_1^A\right)^{n_1-v_1^A}}{\binom{n_1}{v_1^A}\left(1-\gamma_1^B\right)^{v_1^A}\left(\gamma_1^B\right)^{n_1-v_1^A}} > \frac{1-q_2}{q_2} \\
&\Leftrightarrow v_1^A \ln\left(\frac{\gamma_1^A}{1-\gamma_1^B}\right) + \left(n_1 - v_1^A\right)\ln\left(\frac{1-\gamma_1^A}{\gamma_1^B}\right) > \ln\left(\frac{1-q_2}{q_2}\right) \\
&\Leftrightarrow v_1^A\left(\ln\left(\frac{\gamma_1^A}{1-\gamma_1^B}\right) + \ln\left(\frac{\gamma_1^B}{1-\gamma_1^A}\right)\right) > n_1 \ln\left(\frac{\gamma_1^B}{1-\gamma_1^A}\right) - \ln\left(\frac{q_2}{1-q_2}\right) \\
&\Leftrightarrow v_1^A > \Omega^B(n_1) = \frac{n_1 \ln\left(\frac{\gamma_1^B}{1-\gamma_1^A}\right) - \ln\left(\frac{q_2}{1-q_2}\right)}{\ln\left(\frac{\gamma_1^A}{1-\gamma_1^B}\right) + \ln\left(\frac{\gamma_1^B}{1-\gamma_1^A}\right)}.
\end{aligned}$$

and will vote for candidate B otherwise. Symmetrically, an uncommitted voter receiving a signal of $\theta_{i,2} = B$ will vote for candidate A if and only if

$$\begin{aligned}
&\Leftrightarrow \Pr\left\{v_1^A \mid c^{HQ} = A\right\}\underbrace{\Pr\left\{\theta_{i,2} = B \mid c^{HQ} = A\right\}}_{1-q_2} > \Pr\left\{v_1^A \mid c^{HQ} = B\right\}\underbrace{\Pr\left\{\theta_{i,2} = B \mid c^{HQ} = B\right\}}_{q_2} \\
&\Leftrightarrow v_1^A > \Omega^A(n_1) = \frac{n_1 \ln\left(\frac{\gamma_1^B}{1-\gamma_1^A}\right) + \ln\left(\frac{q_2}{1-q_2}\right)}{\ln\left(\frac{\gamma_1^A}{1-\gamma_1^B}\right) + \ln\left(\frac{\gamma_1^B}{1-\gamma_1^A}\right)}.
\end{aligned}$$

and will vote for B otherwise. It follows that the optimal voting decision for a second period uncommitted voter with private signal $\theta_{i,2}$ and signal quality q_2 is given by

$$\mathfrak{v}_{i,2}^* = \begin{cases} A & \text{for } v_1^A > \Omega^A(n_1) \\ \theta_{i,2} & \text{for } \Omega^B(n_1) \leq v_1^A \leq \Omega^A(n_1) \\ B & \text{for } v_1^A < \Omega^B(n_1), \end{cases}$$

■

Proof of Lemma 2. The first period in which all uncommitted voters herd to one candidate can be denoted

$$\widehat{t} \equiv \min\left\{t : \prod\nolimits_{s:\tau(s)<t}\left(\frac{f\left(v_s^A \mid c^{HQ} = A\right)}{f\left(v_s^A \mid c^{HQ} = B\right)}\right) \notin \left[\frac{1-q_{t^{\max}}}{q_{t^{\max}}}, \frac{q_{t^{\max}}}{1-q_{t^{\max}}}\right]\right\},$$

because all voters with q_s such that

$$\prod_{s:\tau(s)<t}\left(\frac{f\left(v_s^A \mid c^{HQ}=A\right)}{f\left(v_s^A \mid c^{HQ}=B\right)}\right)<\frac{1-q_s}{q_s}$$

will herd to candidate B and all voters with q_s such that

$$\prod_{s:\tau(s)<t}\left(\frac{f\left(v_s^A \mid c^{HQ}=A\right)}{f\left(v_s^A \mid c^{HQ}=B\right)}\right)>\frac{q_s}{1-q_s}$$

will herd to candidate A, as shown in lemma 1. Given the assumption $q_s \geq q_{s'}$ for all s, s' such that $\tau(s)<\tau(s')$, all uncommitted voters in any period $t>\widehat{t}$ have signal qualities $q_{s'} \leq q_{\widehat{t}^{\max}}$. This means that

$$\begin{aligned}&\prod_{s:\tau(s)<t}\left(\frac{f\left(v_s^A \mid c^{HQ}=A\right)}{f\left(v_s^A \mid c^{HQ}=B\right)}\right)<\frac{1-q_{\widehat{t}^{\max}}}{q_{\widehat{t}^{\max}}}\\ \Rightarrow &\prod_{s:\tau(s)<t}\left(\frac{f\left(v_s^A \mid c^{HQ}=A\right)}{f\left(v_s^A \mid c^{HQ}=B\right)}\right)<\frac{1-q_{s'}}{q_{s'}}\end{aligned}$$

and

$$\begin{aligned}&\prod_{s:\tau(s)<t}\left(\frac{f\left(v_s^A \mid c^{HQ}=A\right)}{f\left(v_s^A \mid c^{HQ}=B\right)}\right)>\frac{q_{\widehat{t}^{\max}}}{1-q_{\widehat{t}^{\max}}}\\ \Rightarrow &\prod_{s:\tau(s)<t}\left(\frac{f\left(v_s^A \mid c^{HQ}=A\right)}{f\left(v_s^A \mid c^{HQ}=B\right)}\right)>\frac{q_{s'}}{1-q_{s'}}\end{aligned}$$

so that if uncommitted voters in period $\widehat{t}$ herd to candidate c, then voters in period t would have also herded to candidate c *had they voted* in period $\widehat{t}$. Uncommitted voters in period $\widehat{t}+1$ observe all results from periods $1,\ldots,\widehat{t}-1$ and therefore know that no uncommitted voters in period $\widehat{t}$ voted based on their private information. Therefore, their optimal voting behavior is to vote based only on results from the first $\widehat{t}-1$ states, which given the assumption $q_s \geq q_{s'}$ for all s, s' such that $\tau(s)<\tau(s')$ is to herd to the same candidate that voters in period $\widehat{t}$ herded to. Voters in period $\widehat{t}+2$ will then also know that private information was only revealed in periods $1,\ldots,\widehat{t}-1$ and therefore also herd to the same candidate, and so on for all periods $t>\widehat{t}$. ■

Proof of Theorem 1. Let $p \equiv p^A = p^B$. Also let $V_2(n_1, v_1^A)$ be the probability of electing the better candidate when (n_1, v_1^A) is revealed to the remaining $N-n_1$ voters, and let $V_1(n_1, v_1^A)$ be the probability of electing the better candidate when (n_1, v_1^A) is not revealed. (The subscript refers to the number of effective periods of voting, since not revealing the tally is equivalent to simultaneous voting.) When (i) or (ii) are not satisfied, it is immediate that $V_1(n_1, v_1^A)=V_2(n_1, v_1^A)$, since the failure of (i) means the second period voter will vote according to their signals regardless of learning the tally, and the failure of (ii) means that the winner of the election has already been determined by the current tally. We are left to show that $V_1(n_1, v_1^A)>V_2(n_1, v_1^A)$ for any (q, N, p, n_1, v_1^A) such that (i) and (ii) hold. Let W_k denote the winner of the election when there are k effective periods of voting, so that $V_k(n_1, v_1^A)=\Pr\left\{W_k=c^{HQ}\right\}$. Due to the problem being symmetric when

$p_A = p_B$, we can restrict attention to the case in which $\Omega^A(n_1) < v_1^A < N/2$. In this case, we have that

$$\begin{aligned}
V_k(n_1, v_1^A) &= \Pr\left\{W_k = c^{HQ} \mid \left(n_1, v_1^A\right)\right\} \\
&= \Pr\left\{c^{HQ} = A \mid \left(n_1, v_1^A\right)\right\} \Pr\left\{W_k = A \mid c^{HQ} = A, \left(n_1, v_1^A\right)\right\} \\
&\quad + \Pr\left\{c^{HQ} = B \mid \left(n_1, v_1^A\right)\right\} \Pr\left\{W_k = B \mid c^{HQ} = B, \left(n_1, v_1^A\right)\right\} \qquad (3) \\
&= \frac{\Pr\left\{\left(n_1, v_1^A\right) \mid c^{HQ} = A\right\} \Pr\left(c^{HQ} = A\right)}{\Pr\left\{\left(n_1, v_1^A\right)\right\}} \Pr\left\{W_k = A \mid c^{HQ} = A, \left(n_1, v_1^A\right)\right\} \\
&\quad + \frac{\Pr\left\{\left(n_1, v_1^A\right) \mid \omega = \omega_A\right\} \Pr\left(c^{HQ} = B\right)}{\Pr\left\{\left(n_1, v_1^A\right)\right\}} \Pr\left\{W_k = B \mid c^{HQ} = B, \left(n_1, v_1^A\right)\right\} \\
&= \frac{1}{2\Pr\left\{\left(n_1, v_1^A\right)\right\}} \left[\begin{array}{c} \Pr\left\{\left(n_1, v_1^A\right) \mid c^{HQ} = A\right\} \Pr\left\{W_k = A \mid c^{HQ} = A, \left(n_1, v_1^A\right)\right\} \\ + \Pr\left\{\left(n_1, v_1^A\right) \mid c^{HQ} = B\right\} \Pr\left\{W_k = B \mid c^{HQ} = B, \left(n_1, v_1^A\right)\right\} \end{array}\right]. \qquad (4)
\end{aligned}$$

Because $p^A = p^B$ implies $\gamma^A = \gamma^B$, let $\gamma \equiv \gamma^A = \gamma^B = p + (1-2p)\,q$. Let $f(\cdot; m, \rho)$ and $F(\cdot; m, \rho)$ denote the binomial probability density function and binomial cumulative distribution functions, respectively, for m trials and probability of success ρ. Note that $\Pr\left\{\left(n_1, v_1^A\right) \mid c^{HQ} = A\right\} = f\left(v_1^A; n_1, \gamma\right)$ and $\Pr\left\{\left(n_1, v_1^A\right) \mid c^{HQ} = B\right\} = f\left(v_1^A; n_1, 1-\gamma\right)$, and that $\Pr\left\{\left(n_1, v_1^A\right)\right\} = \frac{1}{2} f\left(v_1^A; n_1, \gamma\right) + \frac{1}{2} f\left(v_1^A; n_1, 1-\gamma\right)$. Furthermore, because we have assumed that $v_1^A > \Omega^A(n_1)$,

$$\Pr\left\{W_k = A \mid c^{HQ} = A, \left(n_1, v_1^A\right)\right\} = \begin{cases} 1 - F\left(\frac{N}{2} - v_1^A; N - n_1, \gamma\right) & \text{for } k = 1, \\ \\ 1 - F\left(\frac{N}{2} - v_1^A; N - n_1, 1-p\right) & \text{for } k = 2 \end{cases}$$

and

$$\Pr\left\{W_k = B \mid c^{HQ} = B, \left(n_1, v_1^A\right)\right\} = \begin{cases} F\left(\frac{N}{2} - v_1^A; N - n_1, 1-\gamma\right) & \text{for } k = 1, \\ \\ F\left(\frac{N}{2} - v_1^A; N - n_1, 1-p\right) & \text{for } k = 2. \end{cases}$$

From equation (4), we have that

$$V_1\left(n_1, v_1^A\right) = \frac{f\left(v_1^A; n_1, \gamma\right)\left(1 - F\left(\frac{N}{2} - v_1^A; N - n_1, \gamma\right)\right) + f\left(v_1^A; n_1, 1-\gamma\right) F\left(\frac{N}{2} - v_1^A; N - n_1, 1-\gamma\right)}{f\left(v_1^A; n_1, \gamma\right) + f\left(v_1^A; n_1, 1-\gamma\right)}$$

and

$$V_2\left(n_1, v_1^A\right) = \frac{f\left(v_1^A; n_1, \gamma\right)\left(1 - F\left(\frac{N}{2} - v_1^A; N - n_1, 1-p\right)\right) + f\left(v_1^A; n_1, 1-\gamma\right) F\left(\frac{N}{2} - v_1^A; N - n_1, 1-p\right)}{f\left(v_1^A; n_1, \gamma\right) + f\left(v_1^A; n_1, 1-\gamma\right)}.$$

Let

$$\begin{aligned}
\Delta\left(q; n_1, v_1^A\right) &\equiv \left[f\left(v_1^A; n_1, \gamma\right) + f\left(v_1^A; n_1, 1-\gamma\right)\right]\left(V_1(n_1, v_1^A) - V_2(n_1, v_1^A)\right) \\
&= \left[f\left(v_1^A; n_1, \gamma\right)\left(1 - F\left(\frac{N}{2} - v_1^A; N - n_1, \gamma\right)\right) + f\left(v_1^A; n_1, 1-\gamma\right) F\left(\frac{N}{2} - v_1^A; N - n_1, 1-\gamma\right)\right] \\
&\quad - \left[f\left(v_1^A; n_1, \gamma\right)\left(1 - F\left(\frac{N}{2} - v_1^A; N - n_1, 1-p\right)\right) + f\left(v_1^A; n_1, 1-\gamma\right) F\left(\frac{N}{2} - v_1^A; N - n_1, 1-p\right)\right] \\
&= f\left(v_1^A; n_1, \gamma\right) \underbrace{\left[F\left(\frac{N}{2} - v_1^A; N - n_1, 1-p\right) - F\left(\frac{N}{2} - v_1^A; N - n_1, \gamma\right)\right]}_{<0} \qquad (5) \\
&\quad + f\left(v_1^A; n_1, 1-\gamma\right) \underbrace{\left[F\left(\frac{N}{2} - v_1^A; N - n_1, 1-\gamma\right) - F\left(\frac{N}{2} - v_1^A; N - n_1, 1-p\right)\right]}_{>0}.
\end{aligned}$$

We want to show that $\Delta\left(q; n_1, v_1^A\right) > 0$ for all $\left(n_1, v_1^A\right)$ such that (i) and (ii) hold. Note that $\Delta\left(q; n_1, v_1^A\right) > 0$ if and only if

$$\begin{aligned}\Phi(q) &\equiv \frac{f\left(v_1^A; n_1, 1-\gamma\right)}{f\left(v_1^A; n_1, \gamma\right)}\left[F\left(\frac{N}{2} - v_1^A; N - n_1, 1-\gamma\right) - F\left(\frac{N}{2} - v_1^A; N - n_1, 1-p\right)\right] \\ &- F\left(\frac{N}{2} - v_1^A; N - n_1, 1-p\right) - F\left(\frac{N}{2} - v_1^A; N - n_1, \gamma\right) > 0.\end{aligned}$$

We can substitute

$$\begin{aligned}\frac{f\left(v_1^A; n_1, 1-\gamma\right)}{f\left(v_1^A; n_1, \gamma\right)} &= \frac{\binom{n_1}{v_1^A}(1-\gamma)^{v_1^A}(\gamma)^{n_1 - v_1^A}}{\binom{n_1}{v_1^A}(\gamma)^{v_1^A}(1-\gamma)^{n_1 - v_1^A}} \\ &= \left(\frac{1-\gamma}{\gamma}\right)^{2v_1^A - n_1},\end{aligned}$$

and so we are left to show that

$$\begin{aligned}\Phi(q) \equiv \left(\frac{1-\gamma}{\gamma}\right)^{2v_1^A - n_1} &\left[F\left(\frac{N}{2} - v_1^A; N - n_1, 1-\gamma\right) - F\left(\frac{N}{2} - v_1^A; N - n_1, 1-p\right)\right] \\ &- F\left(\frac{N}{2} - v_1^A; N - n_1, \gamma\right) - F\left(\frac{N}{2} - v_1^A; N - n_1, 1-p\right) \quad > \quad 0. \qquad (6)\end{aligned}$$

for all $q > 1/2$. (Recall that $\gamma = p + (1-2p)\,q$.) The binomial CDF takes the form

$$\begin{aligned}F\left(\frac{N}{2} - v_1^A; N - n_1, \rho\right) &= \sum_{x=0}^{\frac{N-1}{2} - v_1^A} f(x; N - n_1, \rho) \\ &= \sum_{x=0}^{\frac{N-1}{2} - v_1^A} \binom{N-n_1}{x} (\rho)^x (1-\rho)^{N - n_1 - x},\end{aligned}$$

and so

$$\begin{aligned}F\left(\frac{N}{2} - v_1^A; N - n_1, 1-\gamma\right) - F\left(\frac{N}{2} - v_1^A; N - n_1, 1-p\right) &= \sum_{x=0}^{\frac{N-1}{2} - v_1^A} \binom{N-n_1}{x} (1-\gamma)^x (\gamma)^{N-n_1-x} \\ &- \sum_{x=0}^{\frac{N-1}{2} - v_1^A} \binom{N-n_1}{x} (1-p)^x (p)^{N-n_1-x} \qquad (7) \\ &= \sum_{x=0}^{\frac{N-1}{2} - v_1^A} \binom{N-n_1}{x} \left[\begin{array}{c} (1-\gamma)^x (\gamma)^{N-n_1-x} \\ -(1-p)^x (p)^{N-n_1-x} \end{array}\right] \qquad (8)\end{aligned}$$

and

$$\begin{aligned}F\left(\frac{N}{2} - v_1^A; N - n_1, \gamma\right) - F\left(\frac{N}{2} - v_1^A; N - n_1, 1-p\right) &= \sum_{x=0}^{\frac{N-1}{2} - v_1^A} \binom{N-n_1}{x} (\gamma)^x (1-\gamma)^{N-n_1-x} \\ &- \sum_{x=0}^{\frac{N-1}{2} - v_1^A} \binom{N-n_1}{x} (1-p)^x (p)^{N-n_1-x} \qquad (9) \\ &= \sum_{x=0}^{\frac{N-1}{2} - v_1^A} \binom{N-n_1}{x} \left[\begin{array}{c} (\gamma)^x (1-\gamma)^{N-n_1-x} \\ -(1-p)^x (p)^{N-n_1-x} \end{array}\right] \qquad (10)\end{aligned}$$

Substituting (8) and (10) into (6) yields

$$
\begin{aligned}
\Phi(q) &= \left(\frac{1-\gamma}{\gamma}\right)^{2v_1^A-n_1}\left[\sum_{x=0}^{\frac{N-1}{2}-v_1^A}\binom{N-n_1}{x}\left[(1-\gamma)^x(\gamma)^{N-n_1-x}-(1-p)^x(p)^{N-n_1-x}\right]\right] \\
&\quad -\sum_{x=0}^{\frac{N-1}{2}-v_1^A}\binom{N-n_1}{x}\left[(\gamma)^x(1-\gamma)^{N-n_1-x}-(1-p)^x(p)^{N-n_1-x}\right] \\
&= \sum_{x=0}^{\frac{N-1}{2}-v_1^A}\binom{N-n_1}{x}\left[(1-\gamma)^{2v_1^A-n_1+x}(\gamma)^{N-x-2v_1^A}-\left(\frac{1-\gamma}{\gamma}\right)^{2v_1^A-n_1}(1-p)^x(p)^{N-n_1-x}\right] \\
&\quad -\sum_{x=0}^{\frac{N-1}{2}-v_1^A}\binom{N-n_1}{x}\left[(\gamma)^x(1-\gamma)^{N-n_1-x}-(1-p)^x(p)^{N-n_1-x}\right] \\
&= \sum_{x=0}^{\frac{N-1}{2}-v_1^A}\binom{N-n_1}{x}\left[\begin{array}{c}(1-\gamma)^{2v_1^A-n_1+x}(\gamma)^{N-x-2v_1^A}-\left(\frac{1-\gamma}{\gamma}\right)^{2v_1^A-n_1}(1-p)^x(p)^{N-n_1-x} \\ -(\gamma)^x(1-\gamma)^{N-n_1-x}+(1-p)^x(p)^{N-n_1-x}\end{array}\right] \\
&= \sum_{x=0}^{\frac{N-1}{2}-v_1^A}\binom{N-n_1}{x}\left[\begin{array}{c}(1-\gamma)^{2v_1^A-n_1+x}(\gamma)^{N-x-2v_1^A}-(\gamma)^x(1-\gamma)^{N-n_1-x} \\ +\left(1-\left(\frac{1-\gamma}{\gamma}\right)^{2v_1^A-n_1}\right)(1-p)^x(p)^{N-n_1-x}\end{array}\right] \\
&= \sum_{x=0}^{\frac{N-1}{2}-v_1^A}\binom{N-n_1}{x}\left[\begin{array}{c}\underbrace{(\gamma)^{N-n_1}}_{>0}\underbrace{\left(\left(\frac{1-\gamma}{\gamma}\right)^{2v_1^A-n_1+x}-\left(\frac{1-\gamma}{\gamma}\right)^{N-n_1-x}\right)}_{>0} \\ +\underbrace{\left(1-\left(\frac{1-\gamma}{\gamma}\right)^{2v_1^A-n_1}\right)}_{>0}\underbrace{(1-p)^x(p)^{N-n_1-x}}_{>0}\end{array}\right] \\
&> 0 \text{ for all } q>1/2.
\end{aligned}
$$

for all $x\in\left[0,\frac{N-1}{2}-v_1^A\right]$. The first bracketed term is positive since $\gamma>0$, the second term is positive since $x<\frac{N}{2}-v_1^A \Leftrightarrow 2v_1^A-n_1+x<N-n_1-x$, and $q>1/2 \Leftrightarrow \gamma>1/2 \Leftrightarrow \left(\frac{1-\gamma}{\gamma}\right)<1$. The third term is positive again because $\left(\frac{1-\gamma}{\gamma}\right)<1$ and $2v_1^A-n_1>0$ is implied by $v_1^A>\Omega_A(n_1)$ when $p_A=p_B$. The fourth term is positive since $p\in(0,1)$ is assumed. Because each summand is positive, it is immediate that the sum is positive and hence $\Phi(q)>0$ for all $q>1/2$, which proves the result. ■

Proof of Corollary 1. The party's objective function for a simultaneous election can be rewritten as a weighted average of the probability–assuming all uncommitted voters vote according to their private signals in all periods–of selecting c^{HQ} conditional on each first period result, with each result weighted by its probability of occurrence. Similarly, the party's objective function for a sequential election can be rewritten as a weighted average of the probability–assuming uncommitted voters in future periods herd if it is optimal for them to do so–of selecting c^{HQ} conditional on each first period result, with each result weighted by its probability of occurrence. From Theorem 1, we have that the party is at least as well off not revealing results from the first period to future voters regardless of what that outcome is. This means that the probability of selecting c^{HQ} conditional on any possible first period result is at least as high for a simultaneous election as a sequential one. Given that each component of the weighted averages described above is at least as high for a simultaneous election, it follows immediately that the weighted average itself is at least as high for a simultaneous election. The weighted average is sure to be strictly higher for a simultaneous election from the result in Theorem 1 which states that for parameter values such that $\left(n_1, v_1^A\right)$ does not secure the nomination for either candidate but does induce herding in the second period, revealing the tally results in a strictly lower probability of selecting the higher quality candidate than not revealing. ■

Proof of Theorem 2. The proof of (1) is in two steps. I will show that: (i) it is optimal for the party for the very last voter–assuming she is uncommitted–to herd (necessarily to candidate B, the candidate with fewer loyal voters) rather than use her own private information when she is the only voter not voting in the first period and she is pivotal in determining the election result, and (ii) that conditions (1) and (2) are sufficient to induce the situation described in part (i). To show (i), assume that each candidate has received $\frac{N-1}{2}$ votes after the first $N-1$ voters have voted. Call this event T. The probability that the c^{HQ} will be selected in a simultaneous election is equal to q, the probability that the last voter receives the correct signal. In a sequential election, the probability that c^{HQ} will be selected is equal to

$$
\begin{aligned}
\Pr\left\{c^{HQ}=B \mid T\right\} &= \frac{\Pr\left\{T \mid c^{HQ}=B\right\}\Pr\left\{c^{HQ}=B\right\}}{\Pr\left\{T\right\}} \\
&= \frac{\binom{N-1}{\left(\frac{N-1}{2}\right)}\left(\gamma^B\right)^{\frac{N-1}{2}}\left(1-\gamma^B\right)^{\frac{N-1}{2}}\cdot\left(\frac{1}{2}\right)}{\left(\frac{1}{2}\right)\binom{N-1}{\left(\frac{N-1}{2}\right)}\left(\gamma^B\right)^{\frac{N-1}{2}}\left(1-\gamma^B\right)^{\frac{N-1}{2}}+\left(\frac{1}{2}\right)\binom{N-1}{\left(\frac{N-1}{2}\right)}\left(\gamma^A\right)^{\frac{N-1}{2}}\left(1-\gamma^A\right)^{\frac{N-1}{2}}} \\
&= \frac{\left(\gamma^B\right)^{\frac{N-1}{2}}\left(1-\gamma^B\right)^{\frac{N-1}{2}}}{\left(\gamma^B\right)^{\frac{N-1}{2}}\left(1-\gamma^B\right)^{\frac{N-1}{2}}+\left(\gamma^A\right)^{\frac{N-1}{2}}\left(1-\gamma^A\right)^{\frac{N-1}{2}}} \\
&= \frac{\left[\left(\gamma^B\right)\left(1-\gamma^B\right)\right]^{\frac{N-1}{2}}}{\left[\left(\gamma^B\right)\left(1-\gamma^B\right)\right]^{\frac{N-1}{2}}+\left[\left(\gamma^A\right)\left(1-\gamma^A\right)\right]^{\frac{N-1}{2}}}. \qquad (11)
\end{aligned}
$$

We have that $p^A > p^B \Rightarrow \gamma^A > \gamma^B \Rightarrow \gamma^A\left(1-\gamma^A\right) < \gamma^B\left(1-\gamma^B\right)$, since $\gamma^A > 1/2$ and $\gamma^B > 1/2$. It follows that

$$
\frac{\left[\left(\gamma^A\right)\left(1-\gamma^A\right)\right]^{\frac{N-1}{2}}}{\left[\left(\gamma^B\right)\left(1-\gamma^B\right)\right]^{\frac{N-1}{2}}} < 1 < \frac{1-q}{q}, \qquad (12)
$$

and so from (12) and (11) we have

$$
\begin{aligned}
q &< \frac{1}{1+\frac{\left[\left(\gamma^A\right)\left(1-\gamma^A\right)\right]^{\frac{N-1}{2}}}{\left[\left(\gamma^B\right)\left(1-\gamma^B\right)\right]^{\frac{N-1}{2}}}} \\
&= \frac{\left[\left(\gamma^B\right)\left(1-\gamma^B\right)\right]^{\frac{N-1}{2}}}{\left[\left(\gamma^B\right)\left(1-\gamma^B\right)\right]^{\frac{N-1}{2}}+\left[\left(\gamma^A\right)\left(1-\gamma^A\right)\right]^{\frac{N-1}{2}}} \\
&= \Pr\left\{c^{HQ}=B \mid T\right\},
\end{aligned}
$$

which proves (i). To prove (ii), note that from lemma 2 we have that an uncommitted voter observing event T will herd to candidate B regardless of her private signal if and only if

$$
\frac{\left[\left(\gamma^B\right)\left(1-\gamma^B\right)\right]^{\frac{N-1}{2}}}{\left[\left(\gamma^A\right)\left(1-\gamma^A\right)\right]^{\frac{N-1}{2}}} > \frac{q}{1-q}, \qquad (13)
$$

which is true when $p^A > p^B$ as shown in (12). We can rewrite (13) as

$$
\frac{\gamma^B\left(1-\gamma^B\right)}{\gamma^A\left(1-\gamma^A\right)} > \left(\frac{q}{1-q}\right)^{\frac{2}{N-1}}. \qquad (14)
$$

Note that $RHS\,(14) \searrow 1$ as $N \to \infty$ and that $LHS\,(14) > 1$ for all $p^A > p^B$. Therefore, there exists an N large enough such that (14) holds. Call the value at which (14) holds with equality $N^*(q)$, and (1) follows. To show (2), note that

$$\frac{\gamma^B\left(1-\gamma^B\right)}{\gamma^A\left(1-\gamma^A\right)} = \frac{\left[p^B + \left(1-p^A-p^B\right)q\right]\left[1-p^B-(1-p^A-p^B)q\right]}{\left[p^A + \left(1-p^A-p^B\right)q\right]\left[1-p^A-(1-p^A-p^B)q\right]},$$

and so the limit of $LHS\,(14)$ as $q \searrow 1/2$ is

$$\lim_{q \searrow \frac{1}{2}} \frac{\gamma^B\left(1-\gamma^B\right)}{\gamma^A\left(1-\gamma^A\right)} = 1.$$

while the corresponding limit of $RHS\,(14)$ is also

$$\lim_{q \searrow \frac{1}{2}} \left(\frac{q}{1-q}\right)^{\frac{2}{N-1}} = 1.$$

The relevant factor as to how large N has to be for (14) to occur for an arbitrarily small value of q is how the slopes of $LHS\,(14)$ and $RHS\,(14)$ compare. The value of N must be such that

$$\begin{aligned}
& \left.\frac{\partial LHS\,(14)}{\partial q}\right|_{q=1/2} > \left.\frac{\partial LHS\,(14)}{\partial q}\right|_{q=1/2} \\
\Leftrightarrow\ & \frac{8\left(1-p^A-p^B\right)\left(p^A-p^B\right)}{\left(1+p^A-p^B\right)\left(1-p^A-p^B\right)} > \frac{8}{N-1} \\
\Leftrightarrow\ & N > \frac{1-2\left(p^A\right)^2 + p^A - p^B + 2p^Ap^B}{\left(1-p^A-p^B\right)\left(p^A-p^B\right)},
\end{aligned}$$

which proves (2). ■

References

Ali, S. Nageeb, & Kartik, Navin. 2006. Social Learning in Elections. *working paper, University of California, San Diego.*

Banerjee, Abhijit V. 1992. A Simple Model of Herd Behavior. *The Quarterly Journal of Economics*, **107**(3), 797–817.

Bartels, Larry M. 1988. *Presidential Primaries and the Dynamics of Public Choice.* Princeton, NJ: Princeton University Press.

Bikhchandani, Sushil, Hirshleifer, David, & Welch, Ivo. 1992. A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Poltical Economy*, **100**(5), 992–1026.

Callander, Steven. 2007. Bandwagons and Momentum in Sequential Voting. *Review of Economic Studies*, **74**(3), 653–684.

Campbell, Angus, Converse, Philip E., Miller, Warren E., & Stokes, Donald E. 1960. *The American Voter.* Ann Arbor, MI: John Wiley and Sons, Inc.

Dekel, Eddie, & Piccione, Michele. 2000. Sequential Voting Procedures in Symmetric Binary Elections. *Journal of Political Economy*, **108**(1), 34–55.

DiClerico, Robert E., & Davis, James W. 2000. *Choosing Our Choices: Debating the Presidential Nominating Process.* Lanham, MD: Rowman and Littlefield Publishers, Inc.

Fey, Mark. 2000. Informational Cascades and Sequential Voting. *working paper, University of Rochester.*

Goeree, Jacob K., & Groer, Jens. 2007. Welfare reducing polls. *Economic Theory*, **31**(1), 51–68.

Iaryczower, Matias. 2008. Strategic Voting in Sequential Committees. *working paper, California Institute of Technology.*

Klor, Esteban F., & Winter, Eyal. 2007. The welfare effects of public opinion polls. *International Journal of Game Theory*, **35**(3), 379–394.

Klumpp, Tilman, & Polborn, Mattias K. 2006. Primaries and the New Hampshire Effect. *Journal of Public Economics*, **90**(6), 1073–1114.

Knight, Brian, & Schiff, Nathan. 2007. Momentum and Social Learning in Presidential Primaries. *working paper, National Bureau of Economic Research.*

Mayer, William G. (ed). 2004. *The Making of the Presidential Candidates 2004.* Lanham, MD: Rowman and Littlefield Publishers, Inc.

Mayer, William G., & Busch, Andrew W. 2004. *The Front-loading Problem in Presidential Nominations.* Washington, DC: Brookings Institutions Press.

Morgan, David R., & Wilson, Laura Ann. 1990. Diversity in the American States: Updating the Sullivan Index. *The Journal of Federalism*, **20**(Winter), 71–81.

Morton, Rebecca B., & Williams, Kenneth C. 1999. Information Asymmetries and Simultaneous versus Sequential Voting. *American Political Science Review*, **93**(1), 51–67.

Morton, Rebecca B., & Williams, Kenneth C. 2001. *Learning By Voting.* Ann Arbor, MI: The University of Michigan Press.

Palfrey, Thomas R, & Rosenthal, Howard. 1983. A strategic calculus of voting. *Public Choice*, **41**(1), 7–53.

Welch, Ivo. 1992. Sequential Sales, Learning, and Cascades. *The Journal of Finance*, **47**(2), 695–732.

Witt, Jorgen. 1997. Herding Behavior in a Roll-Call Voting Game. *working paper, University of Amsterdam.*